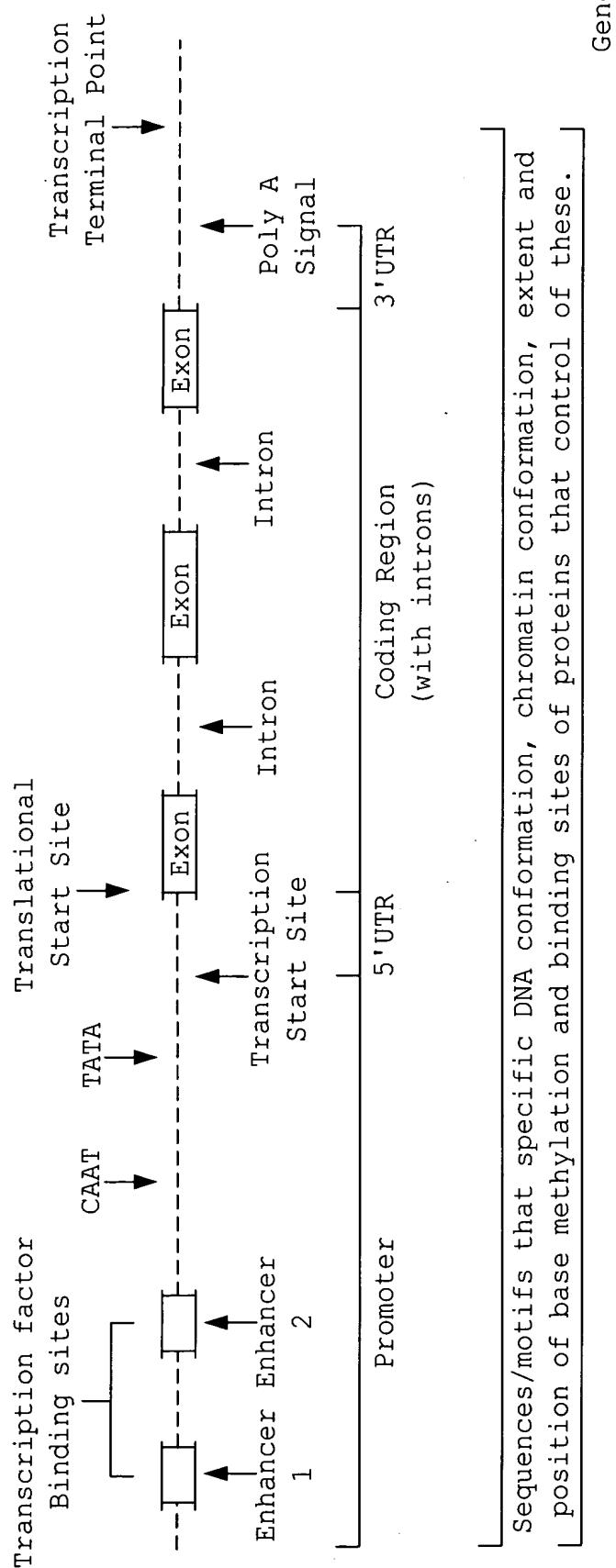




**FIG. 1**

**SCHEMATIC OF A GENE**



## FIG. 2

-4242 GCATGCACTG CCACAAGTAG TGAACTCATG GTTTACCTC CTCAGTAGA  
-4192 AAACCTTTG AGTGAATTG AAGATTATT CTCCAAGAA GGACCCATTG  
-4142 GGCTCCTCC TCTAGGGGG ATAGAACATC AAATTGACTT TATACCGGGG  
-4092 GCAAGCCTAC CAAATAGGCC TCCTTATAGA ACCAACCCCG AGGAAACAAA  
-4042 GGAGATAGAA TCACAAGTTC AAGACTGTGTT GGAGAAGGGT TGGGTTCAAA  
-3992 AGAGCCTAAG CCCTTGTGCT GTACCTGTCT TGTTGGTGCC AAAAAAAAGAT  
-3942 GGAAAATGGC GTATGTGTTG TGATTGTAGA GCAATCAACA ACATCACCAC  
-3892 CAAGTATAGG CATCCAATCC CAAGGCTTGA CGATATGCTT GATGAATTGC  
-3842 ATGGGTCAAC TCTATTCTCC AAAATTGACC TTAAAAGTGG ATATCACCAA  
-3792 ATTCGAATCA AGGAGGGTGA TGAGTGGAAA ACCGCTTTA AGACCAAATT  
-3742 TGGATTATAT GAGTGGTTGG TGATGCCCTT TGGTCTTACT AACGCTCCAA  
-3692 GTACATTCACT GAGGCTTATG AATCACACCT TGAGGGATTG TATAGGTAAA  
-3642 TATGTAGTAG TTTATTGTGA TGATATCTTA GTATATAGTA AAACCCTAGA  
-3592 AGACCATCTA AGTCACCTTA GGGAAAGTTCT TCTAGTTCTT AGGAAAAATA  
-3542 GTCTTTTGC CAATAGGGAT AAGTGTACCT TTTGTGTAGA TAGCGTAGTC  
-3492 TTTTTAGGCT TTATAGTAAA CCAAAAGGGG GTGCATGTAG ATCCCGAGAA  
-3442 AATCAAAGCC ATCCGCGAGT GGCCAACCTC ACAAAATGTA AGTGATGTGA  
-3392 GAAGTTTCA TGGGTTAGCT AGCTTCTATA GAAGGTTGT TCCCAATTTC  
-3342 TCTAGCCTAG CTTCTCCCTT GAATGAACCTT GTAAAAAAAG ATGTTGCATT  
-3292 TTGTTGGAAT GAAAAGCATG AGCAAGCCTT TCAAAGGCTA AAAGCTCACT  
-3242 CACCAATGCA CCCATCCTAT CTCTTCCAAA TTTTCCAAA CTTTGGAGA  
-3192 TAGAGTGTGA TGCATCGGGA GTAGGCATAG TGCGGTTTG TTGCAAGGTG  
-3142 GACACCCCTT GCTTATTCTT GTGAAAAACT CCATGGTGCC ACCCTCACTA  
-3092 CCCCACCTAT GACAAAGACT CTATGCTCTT GTGCGACCCT AAAGACTTGG  
-3042 GGAACACTAC CTTGnGTCCC AAAGAATTG GnTATCCATA GTGATCACGA  
-2992 GTCTTAAAAA TATTTAAAGG GCCAACACAA GCTCAATAAG AGACATGCTA  
-2942 AATGGATGGA ATTTCTTGAA CAATTCCCTT ATGTCATCAA ATACAAGAAA  
-2892 GGGAGCACCA ATATAGTGGC CGATGCTCTT TCTAGACGGC ACACTCTCTT  
-2842 TTCAAAACTA GGTGCCAAA TTCTTGATT TGACCACATA AGAGAGCTTT  
-2792 ATCAAGAAGA TCAAGAACTC TCATCCATCT ATGCCCAATG TCTACATAGA  
-2742 GCACAAGGAG GTTACTATGT GTCCGAGGGG TATCTTTTA AAGAAGGAAA  
-2692 ACTTTGCATT CCCCAAGGAA CACATAGAAA ACTCCTGTGTC AAAGAATCAC  
-2642 ATGAAGGGGG ACTCATGGC CATTTGGAG TTGATAAAAC TCTAGACTTT  
-2592 TAAAAGCAAAT ATTTGTTGG CCACACATGA GGAAAGATGT CCACGACATT  
-2542 GTCTAGAGTA TCTCATGTTT AAAAGCAAAG TCTAGAACAA TGCCGCTGGA  
-2492 CTCTACACCC CTTGCCGAT TGCAAAGCTC CTTGTGAAGA CATTAGCATG  
-2442 GATTCATTT TAGGACTTCC TAGGACTGCA AGAGGCCATG ACTCTATCTT  
-2392 TGTGGTAGTG GACCGTTTA GCAAAATGTC TCACTTTATT CCATGCCACA  
-2342 AAGTAGATGA TGCTAAAAT ATTTCTAAAC TCTTCTTTAG AGAAGTGGTG  
-2292 AGACTCCATG GTCTCCCTAG AAGTATAGTG TCCGATAGAG ATCACCTTAA  
-2242 ATATATAATT ATACACTTGT TTTTTTCTC TTTTTTATT TATCAAGTAA  
-2192 AAAGTATTG TTCTAGATTA TTATGAGTAT ATACTTACTT TCTGTATTTC  
-2142 ATTTCTTCT ATTTTTATG ACGATGAAAT TTCTTATTAT ATCCAGACTT  
-2092 TTCATATATA TTTTATTTC TTTTCCATCT AGATGCTCTG TACTTTCTT  
-2042 CAGTTGAAAT TTCCACTCTC CAACAAAACA TCATTCAAGT TTTGTATAAC  
-1992 ACTGTGACGT TAACCAGTTA AAATAAGAAA ATCATGTAAT ATAAATTATT  
-1942 TCAGTAGATA TTTAGAATT ACAAAATACGA TAAATAATT AATTAAAAAA  
-1892 ATTATTAAC AATGAATTTC TTTGGAAATT AATATAAAAC TTGACTTGT

## FIG. 2 (Continued)

-1792 ACATCGAATT TGGGTGCTTT ATGCCGCTTT ATCTTCATCT GCACCTTCAA  
 -1742 ATTAATAATT TAATTCCGGA AAATAATAAA CCCACACACT GTTTTATGCA  
 -1692 TATATTAAAGA TAAATAAAAG AGAACTATTT TAAAGAATAT AAAATAATAA  
 -1642 ATGTAACAAA TGATGTCACT AAAGAAGAAA AAAATTAACA AGAATTGTAA  
 -1592 TATATTCTT TATGAAATGT TTTGTGCATT ACCGAGAGAG GTCGAACATG  
 -1542 ATACACGCAA GCATCTAACT AGTTTGGTAA TTCCTTTCA ACATCGnTAA  
 -1492 GCACATCACA CTAAAATTAC TTTAAATAGA TAAATTAGAT TCAATTGGAT  
 -1442 GACATTAATT TATAATACTC TATCCAAAAT TATAACTATA AATAAAAAGT  
 -1392 TATTTTACA AAATAAGTAA TGAAAATTAA ATTCTAAAAT TTATAACACT  
 -1342 TTTATGCTGT GTTGTGTTCG AAGCATAGAA AAATAAAAAG TTATTGTTGG  
 -1292 GAATGAAAAG TGAAGAAAAT CATGTAATAA AAACAAAATG ACACGACAAT  
 -1242 CAAAAAAAAG GTTTCATGC AAAACTTTT TCAAAATTAA CACTTTATG  
 -1192 ATGTGTTGTG TTGAAAGTGT AGAAAAACGA AAAGTTATTA TTGGTAATGA  
 -1142 AAAGCGAAGA AAATCACGTA ATAAAAACAA AGCAAGATGG CACGACAATC  
 -1092 AAAAAAAAGT TTCTACACAA AACTTTATTC AAAATTACA ACACTTTAT  
 -1042 GTTGTGTTT GTTCCGAGG TATAGAAAAA CAAAGAATTA GTGTTGGTAA  
 -992 TGAAAAGTGA AGAAAACCAT GTAATGAAAA CAAAATGGCA CGACAATCAA  
 -942 AAAAAAGTTT CACGAAAAT TTTCTTCAAA ATTATATAACA TTTTCATGTT  
 -892 GTGTTGTTT CAAAGCCTAG AAAAACGAAG AGTTACTATT GGTAAATGAAA  
 -842 AGCGAAGAAA ACCACATAAT AAAACAAAAA TGGCACGACA ATCAAGAAAA  
 -792 AGTTTCACA CAAAACTTT TTCAAAATTAA ACTATGTTA TTTCGAAATT  
 -742 TAGAAAAACG AAGAGTTATT ATTAGTAATG AAAAGCGAAG AAAACTACGT  
 -692 AATAAAAAC AAAATGGCAC GACAATAAAA AAAGTTTCA CGCAAAATTT  
 -642 TCTTGGTGC GAGAAAGTTA TATATATTAA TTAATTAAATT TTCATTTACT  
 -592 TTTTCCCTT TTTATTAAAGTAA ATTAAATTAA TTATTATTAA CATTAAAT  
 -542 ATAAATATTA TTTAAATATA AAAATATAA CCTTAATCAA AACAAAGCCT  
 -492 TAATCTAAA TTTACAACAC TTTAACCTT AAAATTAACT TTAAAAGGAA  
 -442 AATGATAGTG TGACAACAA AAAAGTTGTA TACAACCCCTG TCATAGGTT  
 -392 AGAAATAAAAT ATATATAATA AAGAGTAAAT TTGTAATTAA ATGATATAAA  
 -342 AAAGTATTAA AATAATAATA TTTAGAGTAG TAATATGGTT GTATAAAAAA  
 -292 ATGTGGTTGT CCATATATCA TTATTCACTT TAAAATATCA TGACAAATAT  
 -242 TTTCACCGAA AGATGAAAG AACGAAAAGA GCGTTGGATA ATGGAAAAAT  
 -192 ACAAGCAATC TCCCTCCAGT ACTTTGCATA ACATTTGTA TTAGTGATGA  
 -142 GTTTTTATC ATATATATTAGAATATAGG AAAATTAGG AATCACGTGG  
 -92 ATAGCTATAT AATAGTAATA TTTAATTAA TAATGTAGTT GATTTTATTT  
 -42 GTCAACTGGT ATACATAAAAT ATGTGTTGAT AGTGGGTGAC TTGTGGCTTA  
 9 AAGAAATGTC CAGAGGCTGA CAACAACCT GCACAGACTA GCGTAAAC  
 57 **ATG** AAG TCC AAT TTT GCT ATT TTC GTA GTC TTT TCT CTT CTT CTT  
 1 M K S N F A I F V V F S L L L  
 102 CTG GTACCTCTTCAATCTTCTACAAAAACTCTGTTGCTCTTCAACCTCTGTTGTA  
 16 L  
 160 ATTTTGTACACTTTGGAAAATTGAAGCTGATATATATGTAACAAACCTTCAGTTT  
 219 GTCTGCACTGAAACTGATAGAAAAATACGTTGTGGATATATAG GTT GGC  
 17 V G  
 274 AGT TGC AGC TGC GCA AGA AAA GAC ATG AGA GGG TAT TGG AAG GAT  
 19 S C S C A R K D M R G Y W K D  
 319 ATG ATG AAG GAG CAA CCT ATG CCA GAA GCA ATC AAA GAC CTT ATT  
 34 M M K E Q P M P E A I K D L I

**FIG. 2** (Continued)

49    GAG GAT TCA GAA GAA GTG TCA GAA GCA GGG AAG GGT CGT TTT GTT  
      E    D    S    E    E    V    S    E    A    G    K    G    R    F    V  
      AGG GAC TTC GAT GTA AAG CCT AAT GTC ATA TTA TAT CAC ACA CAT  
64    R    D    F    D    V    K    P    N    V    I    L    Y    H    T    H  
      GTT GTG CCC ATG AAG CAG AGG CAG AAG AAT AAA GAT TGA  
79    V    V    P    M    K    Q    R    Q    K    N    K    D    •  
  
493 AGACTATGTGATTGGCAGTTCACTTATTTGGCACCAAATTATGATGCTCTTGTG  
553 TGTTCAAAATTGTACTCAAACCTTGACCCCTTGCAGCATCTTGCTTCTTTGGTCT  
613 TGCTGAATTTCACAGTTACTGTCACGAATAGTTCTCTTCATAATAAGCAACTT  
673 TCCTCTC

### FIG. 3

101001 CAAAACAAAAGCAAATGCCGGTTTCTTATTATTATTCGAACTTAGAC  
100151 CTTTTGTAACGTTCTTAATTTTCTTGATAAAGAACCTATTAT  
100201 ATCTTAGCTAAATATTACCTCATTGTTATGAGCTAAACCACCCAA  
100251 AAATATTGTAGTTGCTTCGGATTAACGCCAAGCAAGTGATTAGAT  
100301 ATATTAAAGGAAAATGAATGAAAGGACAAAAAAATATAACGACAATATT  
100351 TGAATACTGATATTATCTCCATTCTCAAATATTTGATTATTGTGAC  
100401 AATATTGGTTGTTCCATTGCTACATCTTGAGGACATGAAATGATA  
100451 ACATATATATGAACGAGTATAACATTCTCGTTCATTTACAAATAAT  
100501 GTCAATTATGCTAACATTTTATTAAAATTATCCTTATAAGATTTC  
100551 AGTGTATTATTTACCATGGTACTGTAAAGTCGGATGCTATATATATA  
100601 TATATATATATATCAAAATGACACTGAAGAATTATTGAACAAAAA  
100651 CTAAAAACGTAAAATAAAAGAATTTCAAAAATCAAAATTATATA  
100701 AAAATATAGATAAAATGTTAATATAGTACAACCTCTATTCAAACAGAGAG  
100751 AATAAATCTCTATAGACAGTGAATATCCATTATAATAACGAGCAATAGT  
100801 TGTAATGTTGCAGTACAAAAAGAGAATTGTAATATTGTGCATGATTGAG  
100851 AAATCTAAGTTGACTTGAATTAAAGGCTAATTCCAACAAGTACATGTA  
100901 GAAGTTGACTATAGCTATATTTACTACAAATTGATCATTTCAAGAAAG  
100951 ACATTTAAATTAAGATATGCATGCATGACTTGATTGAACCCCCACTCGCTT  
101001 GCTTCGTGCCATTGACAAGATGTTACTTTAAATGCAAGGTAATTATG  
101051 GATATACTCTCTGTATTTTAGTGTAGATATTTCAGAAAATTGTTT  
101101 TTTTCCAAAATCAAATGATATTATTAATTTCATAATAGAATTAAATT  
101151 AATTAAATTAAATTGAAAGATTATGCTGCAGATTAGATTACCATG  
101201 GTGAAATCATGTTAGGTAAATAATAATGATGTTGTTAGGAAAAA  
101251 AAAAATTCTTAATCTTATGTAAGAATGTTAAACTTCATTATAAAAAA  
101301 TATGAAGCAGTATTATATAAGATGTTAACTAATCGAATAATATTTTG  
101351 GGATGAAATTTCATGTTCTGCATATGTTCTAAAAAATAATATGTAAAAATT  
101401 AACATTCAATTGTATGTTATAAGAAATATATGTGAGTTTTGTTAGATAA  
101451 ATAATACCTAAATTAAGAATTGTAAGTTACTGCACCTCAAATATG  
101501 TTATTTTCTTTATTAATAATCAGCAACATTCTAAATGATTTAT  
101551 TTTCTTAAAAATTGAAAAAATGAAATTAGCAAATATGTTAAATTAAA  
101601 ACGAATTAAAGAAAAACTTGAAAGATATGATATGCTTATAAAAAAA  
101651 ACTTGGTGGCGTACCTACTAAATATGATCACATTAGAGATTGATCCTT  
101701 TAGCATATAGTATGTTAGTATGATATCTATATTATTTATTAAAGAGC  
101751 ATATTCATAATATAGTATTATGTTAATTACAATAACGTTCAATTG  
101801 TTATGTTAGTTTAGAAAACCTATTGCGTGTGCATATCAATGTGAGAAA  
101851 GCGACTCCACATGTGAGATGTTGGTCTGAGAAAGCTTCTGCACTGGTC  
101901 GGAACACTTCATGGACTAGAATGCAATCCATCTATTCAAAGAAAAGCAG  
101951 TTGTCCATGCATGCCTCGGTTTTCACATTGGAAGCAGCGCAACAATGT  
102001 CTTACATAATATGCGATCGACTCTGCAACCAATATTCAAGTACATAG  
102051 ACCATGACATCAAAACATTATCACACCGAGAAGAAAGAACGTCAATT  
102101 GGTAACCTTAATGGCATTGCTCGGGTAATTCTCTAAAGAGTTCTCCC  
102151 AAATTTATTGATTCTTGTGTTAACCTTTGCCAAAGAACATACAT  
102201 ATAGATTGACACCATTCAACTTACAAATACAAGTGAATAATAATT  
102251 CAAGCTGAAAGGAATTAAATCATGATCTAAACCTAAACGACAAATTCTT  
102301 CACAAGTGAGAATCACTAATTGACTACCCCTGGTCGCATATACATCATT  
102351 GTTGTAAATCTGAAAATTGGTTGGATTGATCTGATATGTCATTGAT  
102401 AAAACTTGTATTATTAGAATTGCGCAACAGATAAATCATC  
102451 ATCTATTAGAAAATTTCATTGCAACCACAATTACAGGGAAAAGGT

### FIG. 3 (Continued)

102551 ACAAACTTTCAAATACCACTTATGAGAAGCACTAACGATCACCTTTCTT  
102601 TATGACTTTCTTCTAAAGCTAACGCTGGTAGTCATGACTCATGATTATCC  
102651 TTTTCCTAATGGGAATATTGTGGAAGCGGTTCAAATCTTAGACAAAAT  
102701 TCCATGGCCACTAAAAGTTAGCAAAGTTAAAGTTAAAAAAATATG  
102751 AGTGTACTTGGCCATATGCCATTGTTGAGATCATAACAAGAGAAATAA  
102801 TAGTTATTGAAGTTAGATCATAATCACAATACATCATTGCCTTCATCA  
102851 ACATTTCCATGGATTGAGAGGATCAACTCAATACTAATGGTGGGGTC  
102901 TTATTCCATTGCTCTAGCCAATTAAGCAGTTAGGTTATTGTGTA  
102951 CTCTAGTAGTTGCCAAATCAATCTTAATATTCAAATGTTGTAATTCTA  
103001 ATTACGTATAGATAAATGACTAGATAACACGTGGCTTGGTTTATCAGG  
103051 AAAGTTTCCAATCATATATATGAATGTAGAATAGTGTCTTCATTAAT  
103101 TATTAATTAGCATCTCACCATCTGAGACTGGGAGCATGTGACAAGTTGAC  
103151 ATGTGTATTAAGAGAACTTGAGAAAACCACTTTATGATACTCCCATCT  
103201 GAGACTGGGATGAGTACCAATTATAAAAATATGAGTAGTGAAAAAAATAT  
103251 TCAAAAAAAATTCTAACATGTCCTTAAACATTAACTTATAATT  
103301 AACAAACATCTCCAATATGCGTTATGAAAACCTTATAAAACTTTTAT  
103351 AACATGCTTGAAAATTATAAAATCTGTATTAGAAACAAAGTGAT  
103401 ACTTTGAAAATAGACAAATGAAGTGCTATTAGGATATCATA  
103451 AGTCTTAACTGTGGTTGTTGAATTATATACCTGTCAAAATAA  
103501 AACTAAATAAAATTAAATTATTTATAATCATGAAGATAATTATTC  
103551 ATAAAAGATAAAATATAAAATCAACAAATTATTTGTTAATAAAAATAC  
103601 TTTGAGCTCTCTTCATAAGACTTTCCAGCTTCATCTAGAAAATCACA  
103651 TAAATTAAAAGATAAAATAACCGAATAAACATAGTCACATTCAACTCTT  
103701 AGTCTTAGATTGTTAATTTCAAAGGTTAGGTATTGTATATGTTTT  
103751 TTTATTGGGTTGCTAGATTGATCCAAGAAGAAATGACGGGTTGTTAGT  
103801 ATAGATGGTTGTTGAGTTTTCCCCTGGTTACTCGTTGGTTT  
103851 TGTCCCCAGAATTGTCCTGTACTCGCTGGTTATGTCCTACAAAGTCC  
103901 ACGACCATTGCCGGCTTTGTATTCAACTTGAATTCTAAATTGATTG  
103951 ATGAAAAAAATGTATCTCTAAAGTCCATTAGTACCAAAATAACTAT  
104001 ATCATTACTACATAAAATAGTCTGGGTTTCCAAAGTATTGTTGATA  
104051 TATGTTAAGAGTCGAAATAGACACATAGATATAATGTTGAAATG  
104101 TCTCACATAATTATCCTTTCTTCATTCTACCTCTCAAGTTTC  
104151 CAATCCCACCCCTAACGTAATTATTCTAACCTAACGTTAACATTGTTAAC  
104201 AAATCTTAACTAGCTACAAATGTGTATTACAAGTCTAAATAAAACCTA  
104251 CTTAATTCAAAGGTATTAAACCTCTAAATTGATACTTACTTAGTATC  
104301 GATCGGTCTAGTTAGGGTTGGACAAACACCCATCATGGGACGAAATT  
104351 AGTCATTCTACGGTGTCCAAGACACAAATCTGGACTCGATGTGGATATG  
104401 ACACCTCATTATAACTTTAATTCAACTAAACTATTAGGAGGAAG  
104451 AATCGGAATCTGCATATCAATCACAATAGACTATAGTATACTTAGATT  
104501 GATCTAACATGGCTCTCAACTAACGTTAACGTTAACATTGTTAAC  
104551 GAAATCATAAGACATTAAATTAAATCAATGTTCTAAATAACTTGGGT  
104601 TATGTGTCCTCGTAGAGCTAATGTGCACACACAATGAAAGTTGACCCGTT  
104651 CACTTGTCCCACTTTATGATCTTCTTTAGGTTAAATCCAACCTTTA  
104701 TAATCTCATCTTGTATCAAACAAAACCTTGGCCTGTCTTTCTAAT  
104751 TAAAGTAACTCTCACGGAGAAAAGCCAACATTCTCTGTTTATT  
104801 TTTTAAGAAAATGAATTCAAGGGGACCCAAATTAAAAGGAAACCA  
104851 AAACTCCTTCTATGTTACTTGAAGTTCTATGTAATCAACAA  
104901 TCCTAACAGTAGAGAATAAAAACATCGTTGGAGGTTTATATTAGC

**FIG. 3** (Continued)

105001 CCTCTGTCAATGGAGCTATATCACTTGTCAATTGCTTAACCCTTGCGG  
105051 GAAGATTGTTATGAAACAGTTAATGGAATTCTAGTTGCCAATGTCACG  
105101 TTTAATATGTTGTCCTATACTTATTGAATCTTATAATCTTGTAT  
105151 AGAATTATCTACTTTAGTATTTACATTAACATAATCTATAGAATTCTT  
105201 CTTGTTCTATACAATTAAACAAGTAATATATTCTAATACATATTAAAA  
105251 ATGGTGGTGTGCTATCTGAGCTGTAATAGTTGATTGCTCCAGAGAAGAA  
105301 TAGACAAAAATCCTACTTAAGAGGCCACCCTGAAAATTAGACAA  
105351 GAAAAATTAAACAAAATTAGGTTACACATATTATCATTATATATGCA  
105401 CAACACAAAGTTGACCTTGCAATGTACTATTGAATAAAATAATGC  
105451 AAGAAGAGAGGAATTATCACTGTTACCAAGAAAACAACCTCCTCTAAAC  
105501 AGGTCTCTATATATATAACCTTAACACCTAAAGAATTAAACACAGATCAA  
105551 GAAAAAATCCTCAAAACAAAAGTTAAAGCAGAC **ATG** AAG CAA CAG CAA  
1 M K Q Q Q  
105599 CGT TAC TTG GTC GTC TTC ATC GTC CTT TTA AGC TTT CTT  
6 R Y L V V F I V L L S F L  
105638 CTG GTAAAGCTTCTTCCTTAATTATTAAAAACCCTAATTAAAGATCTCATATA  
19 L  
105691 TCTGAATGTTGTATATATTGTTGGTATAG TTT GTG AAT CTG AGT  
20 F V N L S  
105736 GAA GGA AGA ACA GGA GGA GTT GCA GAA GAA TAT TGG AAG  
25 E G R T G G V A E E Y W K  
105775 AAG ATG ATG AAG AAT GAA CCG TTG CCT GAA CCA ATC AAA  
38 K M M K N E P L P E P I K  
105814 GAG CTT CTC AAC AAT CCT TTT AGG ACC GCA CAA GAG AGA  
51 E L L N N P F R T A Q E R  
105853 TTC ATC CAG AAT TTC GAC ACC AAA TCT GTT GTC ATC ATC  
64 F I Q N F D T K S V V I I  
105892 TAC CAC AAT CCT AAT GAA TAA TCAATGAAGTCTCTCATATAG  
77 Y H N P N E •  
105934 ATATCTATGACTTTAATTGTTATGTATGGATCGACTTACGTGCA  
105984 CGTATATGTTATTAATTAAAGAAAAGCTGCTTGAGTTGTGTGTT  
106034 ATACACGTATACTAAATATGTTCTGTTAGTGCAGAAATGTTAACCTAG  
106094 CTATAAGGGATTTTGTCTTTTTGTTACCATTAATGTGAGTGA  
106144 GTGAGTTTGTGTGATGAAAATTAGATTGCTCACATTGTTGATA  
106194 TATATAAATCAATATACTGTGCCCTTCGTGCTGTTCTTATATTATT  
106244 TGTGACATTAATTATCTTATCAAAAATTATTAAACTGTGTC  
106294 CTATGGAAAAGATGAACAATATGAGTTAACCTCATCTCAAGGAGATTCT  
106344 TTTTGTGTTGTTTTC

## FIG. 4

1 AAGCTTACAAATGTCCCCAAAGATGAAACCACGTTATTATTAGTAAATCCTGAAAAGG  
61 TTAACGCTTCTGTCCTCGAATTCTAAACCCTGAAATCTGAAATATCTAGTGGTTAAAATGGAG  
121 ACTTGAGGATATAGTCTCTGAACCAGCTGTCACGGCTGAGTTAGATAACATTACTGAAT  
181 TTCTACGGGAGCGGTTGAAATCACTTCGCCCTTAAGAAGAAGCCTACACCGGGCACC  
241 TTCTTACGCAATTGAAATTAGTCTGCCAGGCAGTCGTGGATCGAAGGTCTTTC  
301 GATACCGAGGAATCTGACTTGCAAGGAATAATTCTAATCACACCACCCAAACCCCTGA  
361 ATACACTTCAGGACCCTCTGAAACCAACTCGTTCGGCTAAATCACAAGAATCTCCCAC  
421 TCATTCCGATTTAGCCAATTAAATATGATATCGGTCTGGGAAGCCGATAAGGAAATTCT  
481 ACAAAAAGAGTTATGAATGAGGAAAATAAGGAAAAGAGAGAACTATTTTAGGTACCC  
541 TGAAAGAGAACGAGAAAATTAGAAAAAAACTACTCTCATCTGTACACTGTTCAAAA  
601 GAATATCCnnnnnAATGGTTAGATAATATAAGAAAAGGATAAGTATGATTAACGTAAAC  
661 CACGTCGGCAGAAACAAAGTGAATTCCCCCTTAGAGGAAGTCGTTCTTAAATATAAG  
721 AAAACAAAGAAGTAGTCGCCTCCCTTTAAATGATCTCAGAAAACGAGAAGTAAGTA  
781 TAAAAGATATTCAAAATCTACACAGTCAACTAAATTTACTAATCAAATGCTTTCAAT  
841 TAGCAAATAAAAACAAAAGAAAAAGmGAAAATTGAAGAAAATCGTTAATAAAACCAT  
901 TTAAATTCTCAGAAGAAGAGATAAAACAGTAAATGGTCAAACCTTGGATTCTTAT  
961 ACGATGAAGTAAAACAAAAGTTATCTATCTCGTAATAAAAGAAAACCGAAATCTAATA  
1021 ATGATATGCCAAAAGGACAAATCAAATCAAGAAGTTTAGACGAAATCGAAAAGAGAT  
1081 TAAAACAAACTCTGAACGACACAATAATGTGATAGAAGAAACTAAAACCTCAGACTCAT  
1141 GTTCAGAGTCTCCGATCGTATTGAAAAAAATAAAACGTAATAATCAGAGATTCCAGTA  
1201 AGCCGAAATTTCACACTCGCCCCACCTCGATATCATCGAGATGGCGATGGACACCTCA  
1261 GCATTGATGGAATGGATACTGAGTGTATGATGGATGACAGATGATGAATATAGAAAAAC  
1321 TCACGAAATAACAATGGCGCTACAGCATATAGAGTAAAACATACCGAGGAACAAACAAT  
1381 AAAATTAAATTATCTGGATTACGGGAGTATTAAAAGGCTGGTGGGATAATTACCTCAT  
1441 GCCAGAACAAAAGAATTATGTTCTAAGCTGTGTAAAAATAGAAAACGAAGAAGGAATACC  
1501 ACTAATGGTGGAAACATTGGTGGTAGCAATAATTCTATAACTTTATAGGAGATCCAAAGAT  
1561 TTTGAAGAAAGAACATCTTATTACTTCATAATCTAAGATGTCCAACCTTAGGTGACTT  
1621 TAGATGGTATTCTCAGAAAATTCTTAGCTATGGTTTAACAAGGGAAGATTGTAGAGAAC  
1681 TTTCTGGAAAGAACGGTTATAGCTGGATTACGGATATCTTGCTGAAAAGTAAAAGA  
1741 AAATTACAAAAGGAATGCCAACACCCAATTAAAAGATGTACCATACGGGAAATAAG  
1801 TTCAGTTGTAAAAAAATACAGGTCTCAGTTATGCAATAATATGAAAATAGAAAATAAGAT  
1861 AAAAAGAGTGAGAGTCAGGGCATCAAGGAATTAGGGAAATTGTACTCAATACGGTA  
1921 TGAACGAAATACCCCTCCATCAAAAAATAAAAGAAAATAGCAAAAGAACAGGGAG  
1981 AAACAAGCGCTAAACAAGCGCTAACCGACGTAAAATTTAGAAAACGGTTAATT  
2041 TTAGAAAACCATGAAAGTCTAATGATAAGCCCCTATAGTCGTTATAATGTGGACGCA  
2101 TAGGACACATGAAGCGAGACTGTAGACTAAAAGAAAAATTAGTAATTGACCATAAGTG  
2161 ATGAATTAAAAGAACAAATGGAAAACCTCTGATAAAATTCCCTCAGAAGGAGAAC  
2221 GAAGAATCAATAGGAGATTCTGATTACGAAGTATTGGACATGAGGATAACAATTGTAATT  
2281 GTGTCTATAAAATAACGATAAGTAGTGAATTAAAATTGCGTTAGATTGCATTGATA  
2341 AAATTAATAATCCGGAGGAAAAGACCAAGCCTTAATAGACATGAAAAGGCTACTCGTTG  
2401 AAAAAGATGAACCCAGTTCATCTTACAAAAACCTGAATTATAGGATATGTTAAAG  
2461 AAATATTGAGAAAAGCGAAAACATCACATAAGAAATAACCATTAGCGATCTTAATAGTG  
2521 AAATAAATAATTAAAAGCCGAAATCGAATCTATAAAAGTCGAGCTACAAGAATTAAAG  
2581 ATAAAATTATACATGAGGAATCCATCTCCTCTGCCGACGAAAATTACAAGAAGGAGAAG  
2641 CTAGTAGACCTTCCATCAAAGAAATAACATACAAAAGACAAAAGTGGCATGTTAAAG  
2701 CCCTAGAATTGTTGTGACCGTTCATGTGGTCAAAGATGAGTCCTTACCTAA  
2761 CACAATAAAAACGTTACTCTTAAATATCAAAGGAGAGCTACAAATATCAATGAATGAAT  
2821 GACATTAATATTTCTTAAACTTGAATGAGTTGTTTCATAAATATCTGAC

**FIG. 4** (Continued)

2941 ATATTTTATCTCAAAGTCAACGAAATATTATAAAAAGAATCAATTAAAAAAATTATTCT  
3001 TTTGCAGAAAAAAAATTAAAAATATGAAACTCCTCCACACCATTACCATATTATAAA  
3061 TATAAAAAACCTCTCACAAATGTGCATTCTGAAATTCTTATGTTGAGAGATTAATCTC  
3121 TAAAGAAAAAAGGTTGAGAAAGGTGCAGCAACA **ATG** TCT CCA TTC TGT AGA  
1 M S P F C R  
3172 AAC TTT TCA ATG GCA TGG GTG CTT ATG GCA TTT GTG TTG TTT  
7 N F S M A W V L M A F V L F  
3214 GCA AAC AGT GCT ATG CCC ACA AAT GGA TCC ACT GTT GGG GTA  
21 A N S A M P T N G S T V G V  
3256 AAA AAC ATG TTG GGT GGT AAA TTG ATG CTA AAC GTT TTA TGT  
35 K N M L G G K L M L N V L C  
3298 CCC CAT ATT GAT AAG CAA CAC ATT ATC CCG AAT GGT GGT TCA  
49 P H I D K Q H I I P N G G S  
3340 TTT GAG TGG AAG TAC AAT GGT GGT GCT CCA CCA ATA GGA CAA  
63 F E W K Y N G G A P P I G Q  
3382 TCA CCA TTC ATG TGT TTC TTT CGG TGG AAT AAT GTT CAT CAC  
77 S P F M C F F R W N N V H H  
3424 TCC CTT GAT CTG TGT TCA CCA AGC AAG TAT ACT GGT TGT GAA  
91 S L D L C S P S K Y T G C E  
3466 AAT GCC ATT TGG GAA ATC AAA GAA AAG CAA TTT TGT AGG TAC  
105 N A I W E I K E K Q F C R Y  
3508 AGA GGT GGA CCT ATT AAT TAT TTT TGC TAT GAC TGG GAT GAT  
119 R G G P I N Y F C Y D W D D  
3550 TAG TTATATAGATTATTCTATGTTCATCTCAATAAAAATGACTTAGAGTGATTCTT  
3609 AGTTTGCTTAACATTCTACATATTCTAACTATCCGTCACTACCACCGTAACTATAT  
3669 TTATTTAAAATTAGTATCTGTACAGTTATTTTAAAAAGGTATGTGGATTAGAAG  
3729 AGAGATAAAATATGTAGACGGTCACCAACCTTAATTTGAACATATGTAAGACTATATTGA  
3789 CCAAGAATATATGTTAAACTCATTCAATTAAAGACTATATCTCCATTATGATTATGCA  
3849 AATGCAATTAGTTTTTTTCAATTGAAGAATTCAAAAGAAAGTTATCATTAAAAGTAT  
3909 CATTAAATCACTTATATGTTCTTAATATCCTTATTGTTAATAGAATAATTTTTT  
3969 ATCCTTAAATTAAAGTTATTACTACTTTTTTCAATCTTCATTATTTGAAATATT  
4029 TAAAATTATCAATTGGTAACACCCCCAGAAAATACATGTAACATCACTTTTTTTA  
4089 TATTACAAATTATGACTTATAGAAATACAATATTAAAAATATAAGGTTCAAAACTACA  
4149 TCCTAAAGTCTTCAGACCCCTCTGACACATGTATCTGCTCGTATATGTGATACAGTC  
4209 ATCGCAGTTCACAGATAACAGAAAACCAAGGGTAAGCTAATGAAAAAAATTCCATAA  
4269 CATATTAAATTCTATGCAAAAAGAACCGAGCTAAAGTAATCATTATAAACATTCTTAAA  
4329 TATTGTTATATAAAATTCAATATCAATTCTATCATTCAATTAGACCACACATGGATCTA  
4389 TTTCAATCACAAATCATGGATTCTATTAAATCTACTTCGnCTCCAGAAGACTCATT  
4449 AAGTATGCCCTACCAAGAGACTAACACCTAATCAAAGAGAAAATGATCAAGGTAAGTCAA  
4509 ACATCCAATAACGAGTGCTACAGTGGACCCAATGTGTATGAACCTTATCAGCTTCT  
4569 CACCACCTGATATCTTATTCTATATGACGTAGATCATCAGTGAACACTAGAGGATCTCCGT  
4629 TAAACATATGTTTTTATACCTAATGTCATCAAACAACAACTCACACATTATCCCAAATG  
4689 TATGACATCAATTCTATACAATTCTATCATTCAATTATAACATATCATTGAATCACA  
4749 TAACATTAAAAATTCTATACCATTCAAGAACATTCCAAACATCAAAGCAATATTACTT  
4809 TCAAACATCAAAATATAATTATTATTAATAAAGCTt

**FIG. 5**

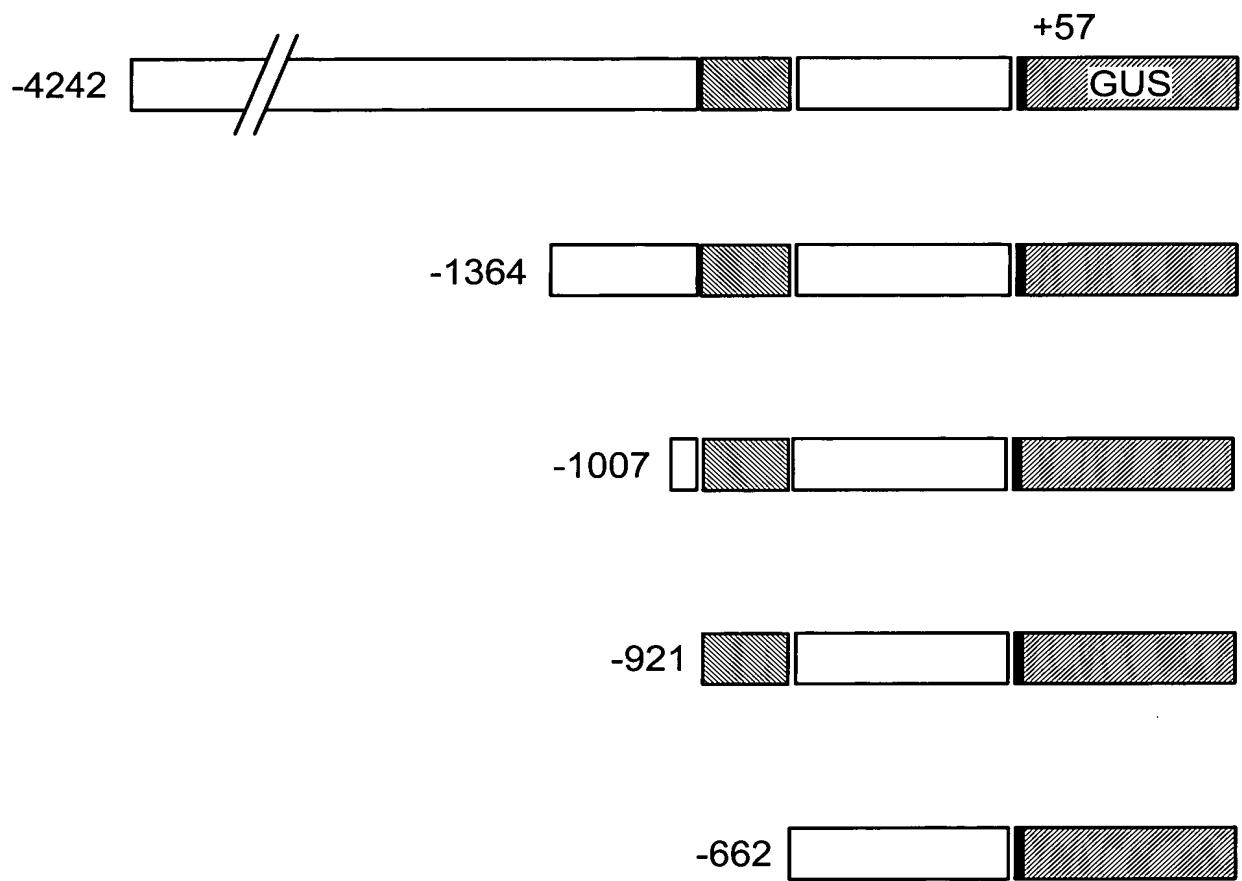
142000 TTATCTTATTCCATATAATTGTTGTTACTTCAAAATTAAATT  
 141950 TTATATTATCTTTACAGTTAAATTAAATAAAATGAAACTTTTTC  
 141900 TTAAATGTGTTAAATATAAAATCAAAAAAGTTGTTATATGGTACATGGC  
 141850 ACAATCTTATAAATTATTAAATTGAAAACGATACTTATATAATAAAATT  
 141800 ATCTTAGTTGACATTTTATTAGTGTTCATCATATTGTTGTTGCTT  
 141750 GATAAGCGTAAAACAAATCAAACCTAACGATACTTATATAATAAAATT  
 141700 TCTTAGTTGACATTTTATTAGTGTCTCAATCATATCTTGTGTTGCTT  
 141650 ATAAGCGTAAAACAAATCAAGTAAAGTTGGGCACCTCAATTGTTTAAAA  
 141600 AAGTTGGGTACCTCAAAATTAAATAGGTCTTGTCAAGATTCTTACAAAAA  
 141550 AAATCTGGAAGAATTATGAAAGAAGGGGGGGAGGGGGGGAGGGGGGG  
 141500 AAGTGAAGATGAATATTCAACAAAAGAGGGTAGGCATGATGTTAAGTGAG  
 141450 TTAAAAAAACTATGTTAATGGAGACAATTCTGTTAACAAACCGTTAAT  
 141400 TGAAAACGATAGCATTCTCTAACATGTAAACGATATTGTTTATC  
 141350 ATAACACTCATTAAATTCTGAGTTCAAAATCATATAAAAGATTAGGGG  
 141300 GGTGTATTCAATTAAAGGATTGAAATGATTTGATTAAAATGACAAATCC  
 141250 CATGTTATTCAACATGAATTGTAACAACTTTAAAATCAAGTGTAA  
 141200 TTAGATTAGTGTATTAAATGTACAACCAACCCACTGTTATTGGAAAC  
 141150 ATTTAAGTAGTGGATTAAAATGACTTGAGTGATTGGGTGGGATTGC  
 141100 AGAAAATTCTTAGTTAAGAATTCAACACATCCAAATCTCATGGTTCAAG  
 141050 TAGAATTGGGAGATTAAATAACAAATCTCCTAACATTACCAAAAGTC  
 141000 CCAAAATCATTAAAAACTCATTAAATTAAATGATTCAAATCTCCAG  
 140950 TTGAATACATCCCCTGGAATTAGAGATTGCTCGATTGGGACCTAAG  
 140900 ATTGAATTGGGGATTAGTTAACATCGTTAACACAAATGACATCGTAT  
 140850 TATTGTTATAGGAAACAATGTCGTTTCAGTTGACATGTATGTTAATAGA  
 140800 AAATTAACTCTATTAAACGGGATTGCTAACCCATTAAACATCGTAACAA  
 140750 ATGGTCAAGTCAATAAAAGTTGGTATTGAAAGTCACGTAAGT  
 140700 TTGATATTATTGAAAAGTCAACATAAATTGATATCTTATTGTTTC  
 140650 GACAGACATAAGGATTACATCAATGTTTAATAAAATTAAAGATTATTA  
 140600 TGACATTTCATTAAATTGCCATTGTTCAACACAAAGATACTC  
 140550 AAAATTGACATACCTAACATCAATTACATCTACATTGTTGACAGCAATTGACGT  
 140500 GCCTGACCACATGGCACATACTGGCAATACATCAATTAAAGGAAAAGG  
 140450 TAGATTGGGATACAATATAATGGAAATAAGTGGAAAGGATCATTGACTAC  
 140400 TTGACTTGTAAACAAACACACAGTATATAACTCATTGACATTACAA  
 140350 ACAACATTGTGCTAGCTAAACTCCCTCCTATTCAAAAAA **ATG**

M

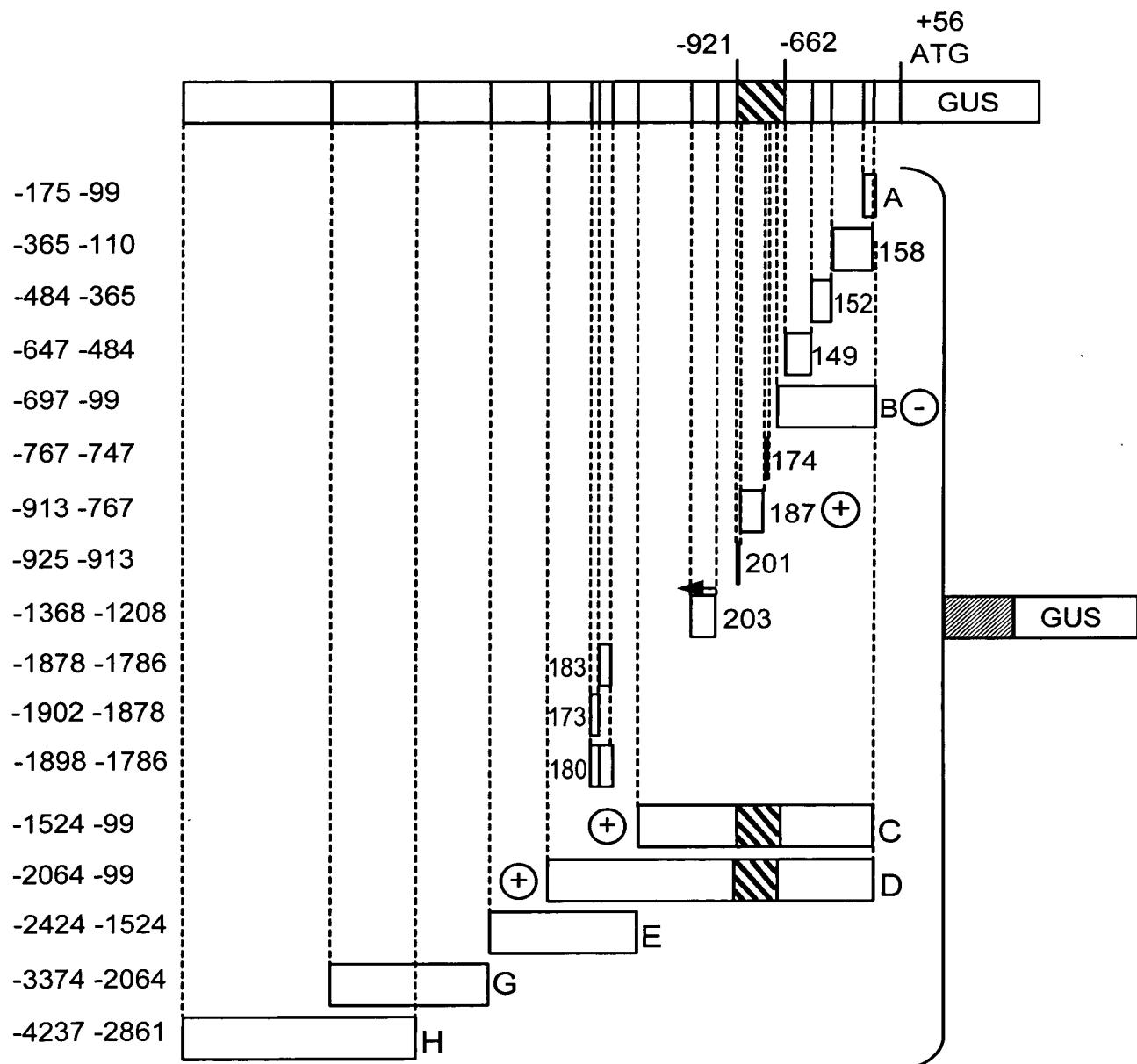
1	GAT	ATT	CCA	AAG	CAA	TAT	CTA	TCA	CTA	TTC	ATA	TTG
2	D	I	P	K	Q	Y	L	S	L	F	I	L
14	ATT	ATC	TTC	ATA	ACT	ACA	AAA	TTA	TCA	CAA	GCC	GAC
14	I	I	F	I	T	T	K	L	S	Q	A	D
26	CAT	AAA	AAC	GAC	ATT	CCA	GTT	CCC	AAC	GAT	CCA	TCA
26	H	K	N	D	I	P	V	P	N	D	P	S
38	TCA	ACA	AAT	TCT	GTG	TTT	CCT	ACC	TCG	AAA	AGA	ACC
38	S	T	N	S	V	F	P	T	S	K	R	T
50	GTG	GAA	ATC	AAT	AAT	GAT	CTC	GGT	AAT	CAG	CTA	ACG
50	V	E	I	N	N	D	L	G	N	Q	L	T
62	TTA	CTG	TAT	CAT	TGT	AAA	TCA	AAA	GAC	GAT	GAT	TTA
62	L	L	Y	H	C	K	S	K	D	D	D	L
140089	GGT	AAC	CGG	ACT	CTG	CAA	CCA	GGT	GAG	TCG	TGG	TCT

**FIG. 5** (Continued)

**FIG. 6**



**FIG. 7**  
**G564 promoter:**  
**Gain of function constructs**



⊕ = Gus Activity in Suspensor

⊖ = No Gus Activity in suspensor

▨ = Region Required for Bus Activity in the Suspensor

## Web Signal Scan Program

Database searched: PLACE

URL: <http://www.dna.affrc.go.jp/btdocs/PLACE/>

This is the sequence you submitted

>G564 promoter (-921 to -662), 450 bases, 3D1AOBF4 checksum.  
 TGAAAAGTGAAGAAAACCATGTAATGAAAACAAAATGGCACGACAATCAA  
 AAAAGTTTCACGAAAATTTCTCAAAATTATAACATTTCATGTT  
 GTGTTTGTTCAAAGCCTAGAAAACGAAGAGTTACTATTGGTAATGAAA  
 AGCGAAGAAAACCACATAATAAAAACAAAATGGCACGACAATCAAGAAA  
 AGTTTCAACACAAAACTTTTCAAAATTACTATGTTATTTCGAAATT  
 TAGAAAAACGAAGAGTTATTAGTAATGAAAAGCGAAGAAAACACGT  
 AATAAAAACAAAATGGCACGACAATAAAAAAGTTTCACGCAAATT  
 TCTTGGTGCGCAGAAAGTTATATATATTAAATTAAATTATTTCAATT  
 TTTTCCCTTTTATTTAAAGTTAAATTATTATTATTTCATTAAAT

Notation: H = A, C, or T

R=A or G

K = G or T

W=A or T

## RESULTS OF YOUR SIGNAL SCAN SEARCH REQUEST

/tmp/signalseqdone.9437: 450 base pairs  
 Signal Database File:

Factor or Site Name	Loc. (Str.)	Signal	Sequence	SITE #	
-300ELEMENT	site	1	(+)	TGHAAARK	S000122
2SSEEDPROTBANAP	site	101	(-)	CAAACAC	S000143
ACGTABOX	site	296	(+)	TACGTA	5000130
ACGTABOX	site	296	(-)	TACGTA	S000130
AP3SV40	site	159	(-)	TGTGGWWW	S000169
CAATBOXI	site	44	(+)	CART	S000028
CAATBOXI	site	189	(+)	CHAT	S000028
CAATBOXI	site	323	(+)	CAAT	S000028
CAATBOXI	site	138	(-)	CHAT	S000028
CANBNNAPA	site	101	(-)	CNAACAC	S000148
CCAATBOXI	site	138	(-)	CCAAT	S000030
CEREGLUBOX2PSLE	site	55	(-)	TGAAAAC	S000033
CEREGLUBOX2PSLE	site	201	(-)	TGAAAAC	S000033
CEREGLUBOX2PSLE	site	333	(-)	TGAAAAC	S000033
DOFCOREZM	site	4	(+)	AAAG	S000265
DOFCOREZM	site	53	(+)	AAAG	S000265
DOFCOREZM	site	112	(+)	AAAG	S000265
DOFCOREZM	site	149	(+)	AAAG	S000265
DOFCOREZM	site	199	(+)	AAAG	S000265
DOFCOREZM	site	282	(+)	AAAG	S000265
DOFCOREZM	site	331	(+)	AAAG	S000265
DOFCOREZM	site	364	(+)	AAAG	S000265
DOFCOREZM	site	419	(+)	AAAG	S000265
DOFCOREZM	site	216	(-)	AAAG	S000265
DOFCOREZM	site	399	(-)	AAAG	S000265
DOFCOREZM	site	408	(-)	AAAG	S000265
GT1CONSENSUS	site	120	(+)	GRWAAW	S000198
GT1CONSENSUS	site	141	(+)	GRWAAW	S000198
GT1CONSENSUS	site	196	(+)	GRWAAW	S000198
GT1CONSENSUS	site	253	(+)	GRWAAW	S000198

Figure 8

GT1CONSENSUS	site	69	( - )	GRWAAW	S000198
GT1CONSENSUS	site	90	( - )	GRWAAW	S000198
GT1CONSENSUS	site	347	( - )	GRWAAW	S000198
GT1CONSENSUS	site	388	( - )	GRWAAW	S000198
GT1CONSENSUS	site	436	( - )	GRWAAW	5000198
GT1CONSENSUS	site	218	( - )	GRWAAW	S000198
GT1CONSENSUS	site	401	( - )	GRWAAW	S000198
GT1CONSENSUS	site	402	( - )	GRWAAW	S000198
MAMMALENHAN	site	158	( - )	GTGGTTTK	S000121
MARTBOX	site	324	( - )	TTWTWTWT	S000067
MRE1	site	356	( - )	TGCRCNC	S000068
NTBBF1ARROLB	site	418	( - )	ACTTTA	S000273
POLASIG1	site	168	( + )	AATAAA	S000080
POLASIG1	site	301	( + )	AATAAA	S000080
POLASIG1	site	324	( + )	AATAAA	S000080
POLASIG1	site	237	( - )	AATAAA	S000080
POLASIG1	site	411	( - )	AATAAA	S000080
POLASIG3	site	268	( - )	AATAAT	S000088
POLASIG3	site	427	( - )	AATAAT	S000088
POLASIG3	site	430	( - )	AATAAT	S000088
POLASIG3	site	433	( - )	AATAAT	S000088
POLLEN1LELAT52	site	11	( + )	AGAAA	S000245
POLLEN1LELAT52	site	119	( + )	AGAAA	S000245
POLLEN1LELAT52	site	156	( + )	AGAAA	S000245
POLLEN1LELAT52	site	195	( + )	AGAAA	S000245
POLLEN1LELAT52	site	252	( + )	AGAAA	S000245
POLLEN1LELAT52	site	289	( + )	AGAAA	S000245
POLLEN1LELAT52	site	362	( + )	AGAAA	S000245
POLLEN1LELAT52	site	71	( - )	AGAAA	S000245
POLLEN1LELAT52	site	349	( - )	AGAAA	S000245
PYRIMIDINEBOXHV	site	400	( + )	TTTTTTCC	S000298
RAV1AAT	site	97	( - )	CAACA	S000314
ROOTMOTIFTAPOX1	site	374	( + )	ATATT	S000098
SEF4MOTIFGM7S	site	170	( - )	RTTTTTR	S000103
SP8BFIBSP8BIB	site	134	( + )	TACTATT	S000184
TATABOX2	site	81	( - )	TATAAAT	S000109
TATABOX3	site	375	( + )	TATTAAT	S000110
TATABOX4	site	368	( - )	TATATAA	S000111
TATABOX5	site	238	( + )	TTATTT	S000203
TATABOX5	site	412	( + )	TTATTT	S000203
TATABOX5	site	434	( + )	TTATTT	S000203

For more information about the SignalScan Program, please contact Dr Dan S. Prestridge Tele: (612) 625-3744 Advanced Biosciences Computing Center. E-mail: damp@biosci.umn.edu 1479 Gortner Ave. University of Minnesota St. Paul, MN 55108 The TFD data is at the gopher site, gopher://genome-gopher.stanford.edu. For more information about the WebSignalScan service, please contact Meena Sakharkar, meena@biomed.nus.sg, BioInformatics centre, NUS .

Database Searched: PlantCARE  
 URL : <http://sphinx.rug.ac.be:8080/PlantCARE/>

Sequence submitted:

>G564 promoter (-921 to -662) 11/21/00

```
+ GAAAAGTGAA GAAAACCATG TAATGAAAAC AAAATGGCAC GACAATCAA AAAAGTTTC ACGCAAATT
+ TTCTTCAAAA TTTATAACAT TTTCATGTTG TGTTGTTTC AAAGCCTAGA AAAACGAAGA GTTACTATTG
+ GTAATGAAAA GCGAAGAAAA CCACATAATA AAAACAAAAT GGCACGACAA TCAAGAAAAA GTTTTCACAC
+ AAAACTTTT TCAAAATTAA CTATGTTAT TTCGAAATT AGAAAAACGA AGAGTTATTA TTAGTAATGA
+ AAAGCGAAGA AAACGTACGTA ATAAAAAACAA AAATGGCACG ACAATAAAA AAGTTTCAC GCAAATTAA
+ CTTGGTGCAC AGAAAGTTAT ATATATTAAT TAATTAATT TCATTTACTT TTTCCCTTT TTATTTAAA
+ GTTAAATTAT TATTATTTTC ATTTAAAAA

- CTTTCACTT CTTTGGTAC ATTACTTTG TTTTACCGTG CTGTTAGTTT TTTGAAAAG TGCCTTTAA
- AAGAAGTTT AAATATTGTA AAAGTACAAC ACAACAAAG TTTCGGATCT TTTGCTTCT CAATGATAAC
- CATTACTTT CGCTTCTTT GGTGTATTAT TTTGTTTA CCGTGCTGTT AGTTCTTTT CAAAAGTGTG
- TTTGAAAAA AGTTTAAAT GATACAAATA AAGCTTAAA TCTTTTGCT TCTCAATAAT AATCATTACT
- TTTCGCTTCT TTTGATGCAT TATTTTTGT TTTACCGTGC TGTTATTTT TTCAAAAGTG CGTTTAA
- GAACCACGCG TCTTCATAATA TATATAATTA ATTAATTAAA AGTAAATGAA AAAAGGGAAA AATAAAATT
- CAATTAAATA ATAATAAAAG TAAATT
```

3-AF1\_binding\_sit

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
3-AF1_binding_sit	ST	260	+	1.000	0.860	AAGAgttatt

Function:

AAGAA-motif

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
AAGAA-motif	Avena sativa	6	+	1.000	0.903	gtgAAGAa
AAGAA-motif	Avena sativa	151	+	1.000	0.870	gcgAAGAa
AAGAA-motif	Avena sativa	284	+	1.000	0.870	gcgAAGAa

Function:

ABRE

Figure 8 (cont'd)

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
ABRE	Hordeum vulgare	293	+	1.000	0.854	actACGtaat
				Function: cis-acting element involved in the abscisic acid responsiveness		

## ACE

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
ACE	Petroselinum crispum	293	+	1.000	0.908	actACGtaat
				Function: cis-acting element involved in light responsiveness		

## AE-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
AE-box	Arabidopsis thaliana	67	-	1.000	0.852	AGAAaatt
AE-box	Arabidopsis thaliana	345	-	1.000	0.852	AGAAaatt
AE-box	Arabidopsis thaliana	361	+	1.000	0.852	AGAAagtt
				Function: part of a module for light response		

## AT1-motif

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
AT1-motif	Solanum tuberosum	409	+	1.000	0.859	ttttATTttaaa
				Function: part of a light responsive module		

## Box\_4

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
Box_4	PC	375	+	1.000	1.000	ATTAat
Box_4	PC	379	+	1.000	1.000	ATTAat
Box_4	PC	383	-	1.000	1.000	ATTAat
				Function:		

## Box\_I

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
Box_I	PS	107	+	1.000	1.000	TTTCaaa
Box_I	PS	203	+	1.000	0.857	TTTCaca
Box_I	PS	219	+	1.000	1.000	TTTCaaa
Box_I	PS	240	+	1.000	0.857	TTTCgaa
Box_I	PS	241	-	1.000	0.857	TTTCgaa
Box_I	PS	249	-	1.000	0.857	TTTCtaa

Function:

**Box\_II**

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
Box_II	ST	139	+	1.000	0.889	TGGTaatga
Box_II	AT	161	+	1.000	0.954	CCACat_aat

Function:

**CAAT-box**

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
CAAT-box	Hordeum vulgare	43	+	1.000	1.000	CAAT
CAAT-box	Arabidopsis thaliana	137	-	1.000	1.000	aCCAAat
CAAT-box	Hordeum vulgare	188	+	1.000	1.000	CAAT
CAAT-box	Hordeum vulgare	322	+	1.000	1.000	CAAT
CAAT-box	Arabidopsis thaliana	351	-	1.000	0.857	aCCAAg

Function: common cis-acting element in promoter and enhancer regions

**ERE**

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
ERE	Dianthus caryophyllus	239	+	1.000	0.875	ATTTcgaa
ERE	Dianthus caryophyllus	241	-	1.000	0.875	ATTTcgaa
ERE	Dianthus caryophyllus	413	+	1.000	0.875	ATTTtaaa
ERE	Dianthus caryophyllus	441	+	1.000	0.875	ATTTtaaaa
ERE	Dianthus caryophyllus	442	-	1.000	0.875	ATTTtaaaa

Function: ethylene-responsive element

**G-box**

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
G-box	Zea mays	17	+	0.842	0.870	CATGta
G-box	Zea mays	38	+	1.000	0.903	CACGac
G-box	Zea mays	94	+	0.842	0.886	CATGtt
G-box	Zea mays	183	+	1.000	0.903	CACGac
G-box	Zea mays	317	+	1.000	0.903	CACGac

Function: cis-acting regulatory element involved in light responsiveness

**GC-repeat**

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
GC-repeat	Oryza sativa	351	-	1.000	1.000	gCACCAag
Function: ?						

## HSE

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
HSE	Brassica oleracea	49	+	0.944	0.878	aAAAAagtt
HSE	Brassica oleracea	50	+	0.944	0.912	aAAAAGttt
HSE	Brassica oleracea	52	-	0.944	0.874	gAAAActtt
HSE	Brassica oleracea	66	-	1.000	0.978	aGAAAattt
HSE	Brassica oleracea	77	-	0.833	0.868	aTAAAtttt
HSE	Brassica oleracea	87	-	1.000	0.853	tGAAAatgt
HSE	Brassica oleracea	196	+	0.944	0.912	aAAAAGttt
HSE	Brassica oleracea	198	-	0.944	0.874	gAAAActtt
HSE	Brassica oleracea	210	+	0.944	0.874	cAAAActtt
HSE	Brassica oleracea	212	-	0.944	0.912	aAAAAGttt
HSE	Brassica oleracea	213	-	0.944	0.878	aAAAagtt
HSE	Brassica oleracea	327	+	0.944	0.878	aAAAAagtt
HSE	Brassica oleracea	328	+	0.944	0.912	aAAAAGttt
HSE	Brassica oleracea	330	-	0.944	0.874	gAAAActtt
HSE	Brassica oleracea	344	-	1.000	0.978	aGAAAattt
HSE	Brassica oleracea	361	+	1.000	0.888	aGAAAgtta

Figure 8 (cont'd)

HSE      *Brassica oleracea*      385      -      1.000      0.853      tGAAAatta

Function: cis-acting element involved in heat stress responsiveness

I-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
I-box	<i>Pisum sativum</i>	93	-	0.857	0.883	aACATga
I-box	<i>Pisum sativum</i>	162	+	0.857	0.883	CACATAaa
I-box	<i>Solanum tuberosum</i>	163	-	1.000	1.000	tATTAtgt
I-box	<i>Pisum sativum</i>	237	-	0.857	0.941	gAAATAaa
I-box	<i>Pisum sativum</i>	367	-	1.000	1.000	tATATAaa
I-box	<i>Pisum sativum</i>	372	+	1.000	0.941	tATATAtt
I-box	<i>Pisum sativum</i>	391	-	0.857	0.941	tAAATga
I-box	<i>Pisum sativum</i>	411	-	0.857	0.883	aAAATAaa
I-box	<i>Pisum sativum</i>	423	+	0.857	0.883	tAAATta
I-box	<i>Solanum tuberosum</i>	424	-	1.000	0.903	aATAAttt
I-box	<i>Arabidopsis thaliana</i>	426	-	1.000	0.863	aATAAtaat
I-box	<i>Arabidopsis thaliana</i>	429	-	1.000	0.863	aATAAtaat
I-box	<i>Solanum tuberosum</i>	431	+	1.000	0.951	tATTAttt
I-box	<i>Pisum sativum</i>	433	-	0.857	0.883	aAAATAaa
I-box	<i>Pisum sativum</i>	439	-	0.857	0.941	tAAATga

Function: part of a light responsive element

P-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
P-box	<i>Oryza sativa</i>	406	+	1.000	0.857	CCTTttt

Function: gibberellin-responsive element

Prolamin\_box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
Prolamin-box	<i>Oryza sativa</i>	145	+	1.000	0.913	tgaAAAGc
Prolamin-box	<i>Oryza sativa</i>	278	+	1.000	0.913	tgaAAAGc

Function: cis-acting regulatory element associated with GCN4

TATA-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
TATA-box	<i>Daucus carota</i>	79	-	1.000	1.000	TATAaatt
TATA-box	<i>Brassica juncea</i>	80	-	1.000	1.000	TATAaat
TATA-box	<i>Helianthus annuus</i>	81	-	1.000	1.000	TATAaaa
TATA-box	<i>Brassica oleracea</i>	82	+	1.000	0.908	tTATAac
TATA-box	<i>Brassica napus</i>	83	-	1.000	0.892	gtTATA
TATA-box	<i>Oryza sativa</i>	117	+	0.818	0.912	TAGAaaa
TATA-box	<i>Oryza sativa</i>	169	+	0.818	0.872	TAAAaac
TATA-box	<i>Zea mays</i>	248	+	0.909	0.879	TTTAgaaa

Figure 8 (cont'd)

TATA-box	Oryza sativa	250	+	0.818	0.912	TAGAaaa
TATA-box	Oryza sativa	302	+	0.818	0.912	TAAAaaa
TATA-box	Oryza sativa	325	+	0.818	0.912	TAAAaaa
TATA-box	Daucus carota	364	-	1.000	0.863	TATAactt
TATA-box	Brassica juncea	365	-	1.000	0.857	TATAact
TATA-box	Zea mays	366	-	1.000	0.879	TATAtaac
TATA-box	Oryza sativa	367	-	1.000	0.956	TATAtaa
TATA-box	Oryza sativa	368	+	1.000	0.929	TATAtat
TATA-box	Oryza sativa	369	-	1.000	0.929	TATAtat
TATA-box	Solanum tuberosum	370	-	1.000	1.000	TATAta
TATA-box	Glycine max	372	+	1.000	0.891	TATAtt
TATA-box	Oryza sativa	407	-	0.818	0.872	TAAAag
TATA-box	Zea mays	413	-	0.909	0.879	TTTAAaat
TATA-box	Zea mays	442	+	0.909	0.879	TTTAAaat

Function: core promoter element around -30 of transcription start

#### TC-rich repeats

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
TC-rich_repeats	NT	7	-	1.000	0.952	gTTTTcttca
TC-rich_repeats	NT	68	+	1.000	1.000	aTTTTcttca
TC-rich_repeats	NT	152	-	1.000	0.909	gTTTTcttcg
TC-rich_repeats	NT	191	-	1.000	0.885	tTTTTcttga
TC-rich_repeats	NT	248	-	1.000	0.914	tTTTTcttaaa
TC-rich_repeats	NT	285	-	1.000	0.909	gTTTTcttcg
TC-rich_repeats	NT	346	+	1.000	0.915	aTTTTcttgg

Function:

#### WUN-motif

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
WUN-motif oleracea	Brassica	18	-	1.000	0.948	tCATTacat
WUN-motif oleracea	Brassica	139	-	1.000	1.000	tCATTacca
WUN-motif oleracea	Brassica	237	+	0.857	0.948	tTATTtcga
WUN-motif oleracea	Brassica	242	-	1.000	1.000	aAATTtcga
WUN-motif oleracea	Brassica	272	-	1.000	0.948	tCATTacta
WUN-motif oleracea	Brassica	296	-	0.857	0.948	tTATTacgt

Function: wound-responsive element

Figure 8 (cont'd)